Section 3

DATA AND DATA COLLECTION

Overview:

Improving your decision making process through the appropriate use of data and learning to collect the right data are crucial Quality skills. By using the tools and techniques in this section, you will be able to gather data that will be meaningful to you and useful in your improvement efforts.

Inside This Section:

What is Data?

Basically, it is the numerical facts and figures which contain the information you will need to form conclusions or make your decisions. Data will generally be presented in descriptive or quantitative form.

Types of Data

MEASUREMENT DATA (Measurements)	Data resulting from a physical measurement Example: Distance, time, weight, etc. (Also known as Measurement or Continuous Data)
ATTRIBUTES DATA (Traits)	Data resulting from a count of units possessing particular characteristics or from a count of the occurrences of those characteristics themselves. (Also called Discrete Data) Example: No. of typos per page, Good / Bad.

Why Collect Data?

- To provide a foundation to "sell" proposed solutions or other actions to those in decision-making positions
- To serve as the basis for timely action (or appropriate non-action)
- To enable you to focus on the real reasons for problems, not just assumptions, symptoms, or "gut" feelings
- To communicate the situation / issues more accurately and effectively
- To allow you to methodically examine the relationship between the occurrence of an event and its cause(s)
- To provide the basis for process control and improvement
- To form a legally valid basis for acceptance or rejection of vendor-supplied items
- To justify or validate opinions or beliefs, even ones long-held or taken for granted.

Five Elements of Useful Data

To be most useful, your data should contain all of these essential elements:

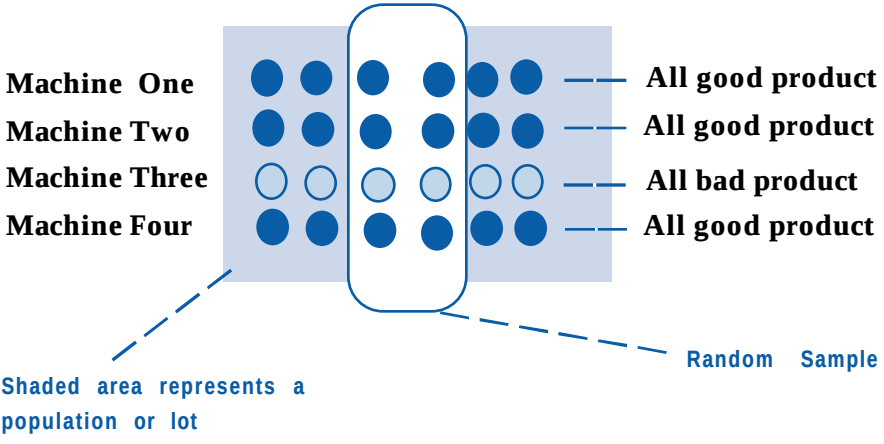
- **Data should be collected in a timely manner**
- **Data should be collected in a consistent manner**
- **Data should be related to your process or the issue being examined**
- **Data should be accurate**
- **Data should be precisely defined (definitions should be agreed-upon by all process participants, i.e.. Supplier/Customer/Process Worker)**

Sampling

Sampling is a technique used to estimate , with a statistical degree of confidence, information concerning a process where measuring or counting the outputs of the entire process is impractical or too expensive. A collection or set of individuals, objects, or measurements whose properties or characteristics are to be analyzed is called a "population." Sampling is merely "surveying" a portion of that population.

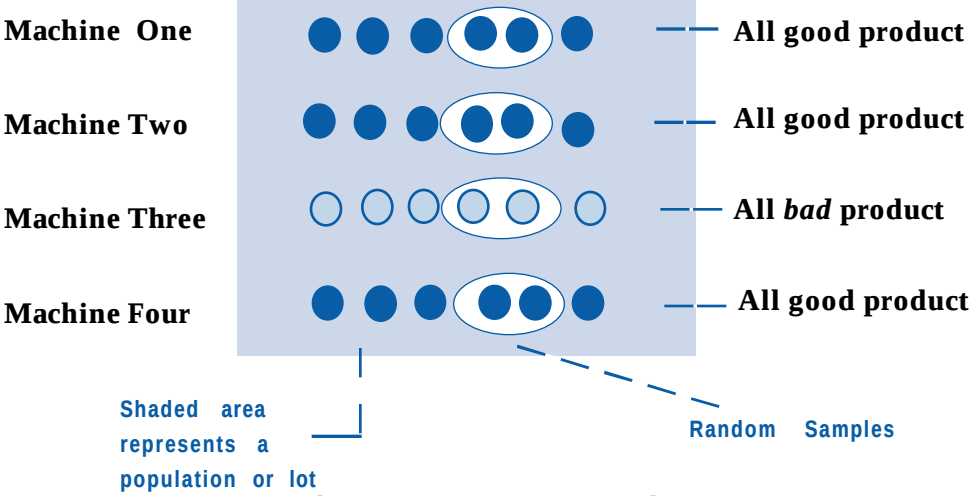
When sampling, it is crucial that a representative portion of the population be selected. A good method for doing this is to use a ***RANDOM SAMPLE***. While a lot could be said here concerning random sampling, the essential point to remember is that **a sample can be described as random if each member of the population had an equal chance of being included in the sample.**

Stratified Random Sampling is similar to regular random sampling, except that the population or lot sampled is divided into subsections (also called strata or layers) which should be as similar as possible. The following will illustrate the advantage of stratified random sampling.



An unstratified random sample could lead us to conclude that our machines are producing 25% BAD product, and that we therefore need tc perform some kind of maintenance or inspection on ALL of them.

To prevent misreading our process this way, we can utilize a stratified random sample. Basically, it would look this way:



As you can see, a random sample comprised of one unit from each of the four layers clearly shows that three machines are producing good product and that only one is not. Only one machine therefore needs to be taken off-line to inspect / perform maintenance on.

Effective Data Collection Strategy

Answering the following eight questions will allow you to develop an effective strategy for collecting data.

WHAT DO WE WANT TO ACCOMPLISH BY COLLECTING DATA?
WHAT DATA IS NEEDED TO ACHIEVE THIS GOAL?
WHERE IN THE PROCESS SHOULD WE COLLECT DATA?
WHAT SAMPLING SCHEME SHOULD WE USE?
HOW MUCH DATA (how many samples/data points) IS NEEDED?
WHEN/HOW LONG SHOULD DATA BE COLLECTED?
HOW WILL WE RECORD THE DATA?
WHO IS RESPONSIBLE FOR COLLECTING THE DATA?

Once you have developed answers to these questions, review your data collection plan for consistency, completeness, and the potential for gaining commitment to your data collection effort from others involved in the process being observed.

CHECKSHEET

What it is:

A check sheet is a simple form you can use to collect data in an organized manner and easily convert it into readily useful information. With a check sheet, you can:

- Collect data with minimal effort.
- Convert raw data into useful information.
- Translate opinions of what is happening into what is actually happening. In other words, "I think the problem is . . ." becomes "The data says the problem is . . ."

How to use it:

Clearly identify what is being observed. The events being observed should be clearly labeled. Everyone has to be looking for the same thing.

Keep the data collection process as easy as possible. Collecting data should not become a job in and of itself. Simple check marks are easiest.

Group the data. Collected data should be grouped in a way that makes the data valuable and reliable. Similar problems must be in similar groups.

Be creative. Try to create a format that will give you the most information with the least amount of effort.

Tabular Check Sheet Example

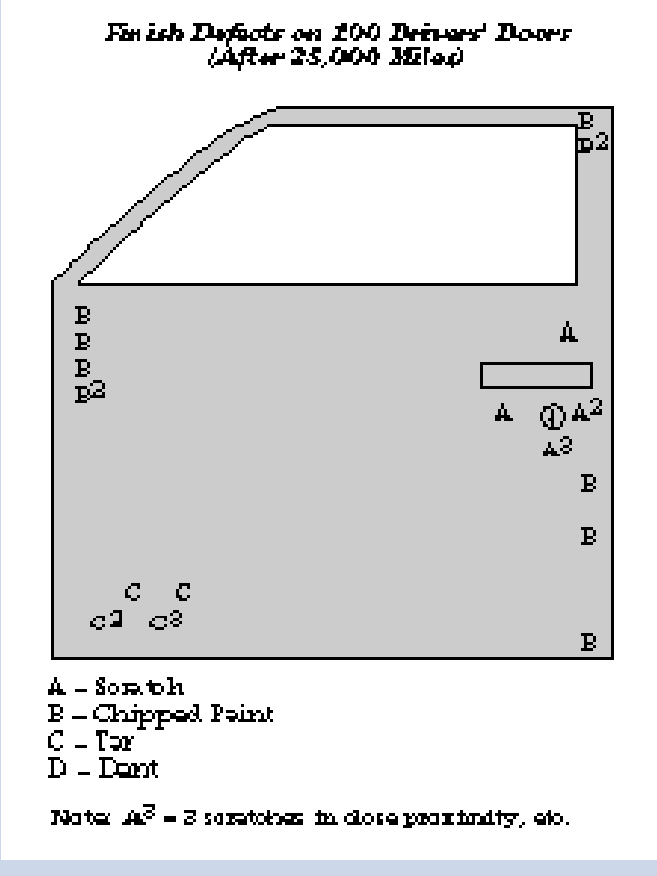
Reasons for Misplaced Letters

Defect	Map 6	Map 7	Map 8	Total Defects
Wrong Mailbox	III	IIII	IIII	11
Wrong City	II	𝍸	𝍸 I	13
Wrong Zip Code	𝍸 IIII	𝍸 II	𝍸	21
Old Office Symbol	III	I	III	7
Total Defects	17	17	16	52

Pictorial Check Sheet v.s. Tabular Check Sheet Example

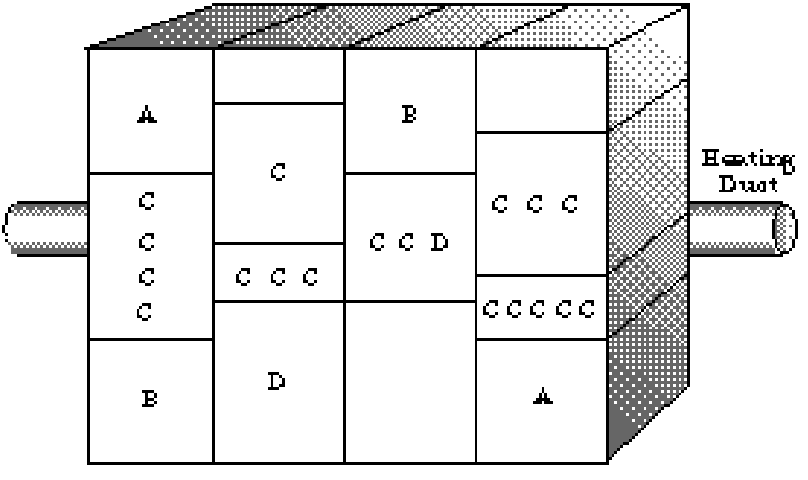
This example shows how a pictorial check sheet can give you much more information than a tabular check sheet.

Defect	Tally	Total
Scratch	𝍸 ||	7
Chipped Paint	𝍸 𝍸 |	11
Tar	𝍸 ||	7
Dent		0
Total		25



Pictorial Check Sheet Example

During testing, the control console in a particular electronic component experienced an unusually high failure rate in some of its black boxes. To help analyze the failures, the managers in the program office developed a pictorial check sheet. After 120 days of testing, the following check sheet showed the types of failures associated with each component.



The managers noticed that most of the failures (18 of 24) occurred in boxes along the center of the equipment racks, and that most of these failures were type C, shutdown during test. An inspection of the facility revealed that a heating duct ran directly behind these boxes, and the resulting high temperatures caused the equipment to overheat and fail.